Experimental Identification of the Use of Hedges in the Simplification of Numerical Expressions

Susana Bautista and **Raquel Hervás** and **Pablo Gervás**
Universidad Complutense de Madrid, Spain
{raquelhb,subautis}@fdi.ucm.es, pgervas@sip.ucm.es

Richard Power and **Sandra Williams**
Department of Computing, The Open University, Milton Keynes MK76AA, UK
{r.power,s.h.williams}@open.ac.uk

Abstract

Numerical information is very common in all kinds of documents from newspapers and magazines to household bills and wage slips. However, many people find it difficult to understand, particularly people with poor education and disabilities. Sometimes numerical information is presented with hedges that modify the meaning. A numerical hedge is a word or phrase employed to indicate explicitly that some loss of precision has taken place (e.g., "around") and it may also indicate the direction of approximation (e.g., "more than"). This paper presents a study of the use of numerical hedges that is part of research investigating the process of rewriting difficult numerical expressions in simpler ways. We carried out a survey in which experts in numeracy were asked to simplify a range of proportion expressions and analysed the results to obtain guidelines for automating the simplification task.

1 Introduction

All public information services and documents should be accessible in such a way that makes them easily understood by everybody, according to the United Nations (1994). Nowadays, a large percentage of information expressed in daily news comes in the form of numerical expressions (statistics of economy, demography data, etc). But many people have problems with understanding such expressions -e.g., people with limited education or some kind of mental disability.

Lack of ability to understand numerical information is an even greater problem than poor literacy. A U.K. Government Survey in 2003 estimated that 6.8 million adults had insufficient numeracy skills to perform simple everyday tasks such as paying house-hold bills and understanding wage slips, and 23.8 million adults would be unable to achieve grade C in the GCSE maths examination for 16 year-old school children (Williams et al., 2003).

A first possible approach to solve this important social problem is making numerical information accessible by rewriting difficult numerical expressions using alternative wordings that are easier to understand. Some loss of precision could have positive advantages for numerate people as well as less numerate. Such an approach would require a set of rewriting strategies yielding expressions that are linguistically correct, easier to understand than the original, and as close as possible to the original meaning.

In rewriting, hedges play an important role. For example,"50.9%" could be rewritten as "just over half" using the hedge "just over". In this kind of simplification, hedges indicate that the original number has been approximated and, in some cases, also the direction of approximation.

This paper presents a preliminary study of the use of hedges when numerical expressions are simplified to make them more accessible. We have carried out a survey in which experts in numeracy were asked to simplify a range of proportion expressions to obtain guidelines for developing the numerical expressions simplification task automatically. As a first step towards more complex simplification strategies, we

Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies, pages 128–136,
Edinburgh, Scotland, UK, July 30, 2011. ©2011 Association for Computational Linguistics

are trying to simplify numerical expressions without losing substantial information. Our study does not have a particular kind of disability in mind. Rather, we aim to simplify according to levels of difficulty defined in the Mathematics Curriculum of the Qualifications and Curriculum Authority (1999). Adaptation to particular types of users is beyond the scope of this paper.

2 Background

Text simplification, a relative new task in Natural Language Processing, has been directed mainly at syntactic constructions and lexical choices that some readers find difficult, such as long sentences, passives, coordinate and subordinate clauses, abstract words, low frequency words, and abbreviations. Chandrasekar et al. (1996) introduced a two-stage process, first transforming from sentence to syntactic tree, then from syntactic tree to new sentence; Siddharthan (2002) instead proposed a three-stage process comprising analysis, transformation and generation. In 1998, the project PSET (Carroll et al., 1998) employed lexical as well as syntactic simplifications. Other researchers have focused on the generation of readable texts for readers with low basic skills (Williams and Reiter, 2005), and for teaching foreign languages (Petersen and Ostendorf, 2007). There has been some previous work on numerical expressions but more for experts than for people who have difficulties with numeracy (Ellen Peters and Dieckmann, 2007), (Nathan F. Dieckmann and Peters, 2009), (Ann M. Bisantz and Munch, 2005), (Mishra H, 2011). However, to our knowledge, there have been no previous attempts to automatically simplify *numerical* information in texts.

A corpus of numerical expressions was collected for the NUMGEN project (Williams and Power, 2009). The corpus contains 10 sets of newspaper articles and scientific papers (110 texts in total). Each set is a collection of articles on the same topic — e.g., the increased risk of breast cancer in red meat eaters, and the decline in the puffin population on the Isle of May. Within each set, identical numerical facts are presented in a variety of linguistic and mathematical forms.

3 Experiment

Our survey took the form of a questionnaire in which participants were shown a sentence containing one or more numerical expressions which they were asked to simplify using hedges if necessary.

3.1 Materials

Our simplification strategies are focused at two levels: decimal percentages and whole-number percentages. For the survey we chose three sets of candidate sentences from the NUMGEN corpus: eight sentences containing only decimal percentages and two sets of eight sentences containing mixed whole-number and decimal percentages. The number of numerical expressions are more than eight because some sentences contained more than one proportion expression.

A wide spread of proportion values was present in each set, including the two end points at nearly 0.0 and almost 1.0. We also included some numerical expressions with hedges and sentences from different topics in the corpus. In short, we included as many variations in context, precision and different wordings as possible.

3.2 Participants

We carried out the survey with primary or secondary school mathematics teachers or adult basic numeracy tutors, all native English speakers. We found them through personal contacts and posts to Internet forums. The task of simplifying numerical expressions is difficult, but it is a task that this group seemed well qualified to tackle since they are highly numerate and accustomed to talking to people who do not understand mathematical concepts very well. Our experimental evaluation involved 34 participants who answered at least one question in our survey (some participants did not complete it).

3.3 Survey Design and Implementation

The survey was divided into three parts as follows:

1. Simplification of numerical expressions for a person who can not understand percentages
2. Simplification of numerical expressions for a person who can not understand decimals

3. Free simplification of numerical expressions for a person with poor numeracy

Each part of the survey is considered as a different kind of simplification: (1) simplification with no percentages, (2) simplification with no decimals and (3) free simplification.

For part (2), the set of sentences containing only decimal percentages was used. One of the two mixed sets of sentences with whole-number and decimal percentages was used for part (1) and the other for part (3). The experiment was presented on SurveyMonkey[1], a commonly-used provider of web surveys. The survey was configured so that participants could leave the questionnaire and later continue with it.

We asked participants to provide simplifications for numerical expressions that were marked by square brackets in each sentence. Below the sentence, each bracketed number was shown beside a text box in which the participant was asked to type the simplified version. Our instructions said that numerical expressions could be simplified using any format: number words, digits, fractions, ratios, etc. and that hedges such as 'more than', 'almost' and so on could be introduced if necessary. Participants were also told that the meaning of the simplified expression should be as close to the original expression as possible and that, if necessary, they could rewrite part of the original sentence. Figure 1 shows a screenshot of part of the questionnaire.

3.4 Underlying assumptions

A numerical expression (NE) is considered to be a phrase that represents a quantity, sometimes modified by a numerical hedge as in "less than a quarter" or "about 20%". We have restricted coverage to proportions -i.e., fractions, ratios and percentages. We had five hypotheses:

- **H1:** The use of hedges to accompany the simplified numerical expression is influenced by the simplification strategy selected. We consider the use of fractions, ratios and percentages like simplification strategies.
- **H2:** The use of hedges to simplify the numerical expression is influenced by the value of the proportion, with values in the central range (say 0.2 to 0.8) and values at the extreme ranges (say 0.0-0.2 and 0.8-1.0) having a different use of hedges.
- **H3:** The loss of precision allowed for the simplified numerical expression is influenced by the simplification strategy selected.
- **H4:** There is some kind of correlation between the loss of precision and the use of hedges, in such a way that the increase or decrease in the former influences changes in the latter.
- **H5:** As an specific case of H4, when writers choose numerical expressions for readers with low numeracy, they do not tend to use hedges if they are not losing precision.

4 Results

The results of the survey were carefully analyzed as follows. First, within each block of questions, a set of simplification strategies was identified for each specific numerical expression. These strategies were then grouped together according to the mathematical forms and/or linguistic expressions employed (fractions, ratios, percentages).

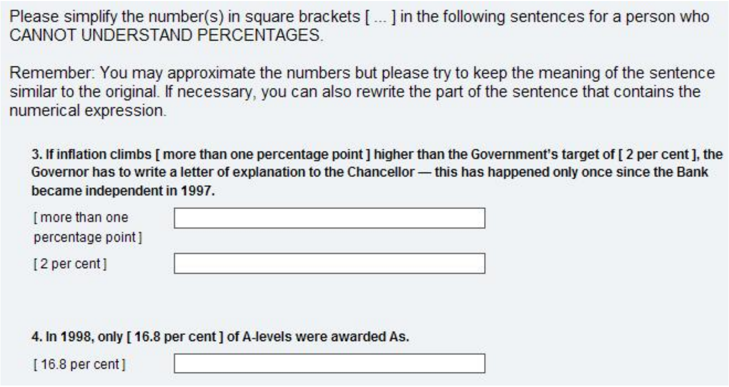
With a view to using these data to design an automated simplification system, these data have to be analyzed in terms of pairs of a given input numerical expression and the simplified expression resulting from applying a specific simplification strategy. For such pairings, three important features must be considered as relevant to choosing a realization:

- Whether any numbers in the expression are realized as one of the different types of available expressions (fractions, ratios, percentages).
- The loss of precision involved in the simplification.
- The possible use of a hedge to cover this loss of precision explicitly in the simplified expression.

To calculate the loss of precision, we defined Equation 1.

$$error = \frac{(simplifiedNE - originalNE)}{originalNE} \quad (1)$$

[1] www.surveymonkey.com



Please simplify the number(s) in square brackets [...] in the following sentences for a person who CANNOT UNDERSTAND PERCENTAGES.

Remember: You may approximate the numbers but please try to keep the meaning of the sentence similar to the original. If necessary, you can also rewrite the part of the sentence that contains the numerical expression.

3. If inflation climbs [more than one percentage point] higher than the Government's target of [2 per cent], the Governor has to write a letter of explanation to the Chancellor — this has happened only once since the Bank became independent in 1997.

[more than one percentage point]

[2 per cent]

4. In 1998, only [16.8 per cent] of A-levels were awarded As.

[16.8 per cent]

Figure 1: Screenshot of part of the questionnaire.

The set of pairings of input expression and observed simplification strategies, loss of precision and use of hedges as found in the results of the survey is given in Tables 1, 2 and 3. For each input numerical expression, the set of available simplification strategies is represented as three lines in the table. For each pairing, three columns are shown in the table. Empty cells represent that the strategy was not used. The first column presents the relative frequency of usage with respect to the total set of possible simplification strategies used for that expression. The second column captures the loss of precision involved, represented in terms of the ratio between the value of the difference between the original numerical value in the input expression and the numerical value that is conveyed by the corresponding simplified expression (using Equation 1). This ratio is also expressed as a percentage. The third column indicates the percentage of simplified numerical expressions that contained a hedge. All of them are mean values.

Each line represents one kind of simplification strategy used to simplify the original numerical expression. Another point to explain is that frequencies that belong to the same expression do not always add up to 100%. This is because a small number of others kinds of simplification strategies, like deletions or rewriting of the whole sentence, are not shown in the table. Moreover, we must keep in mind that not all participants answered each question of the survey.

Table 1 presents the relationships identified between the original numerical expressions and the simplification strategies (presented as lines) for the results of the first part of the survey (simplification of numerical expressions for a person who can not understand percentages). All the values are represented in percentages. Table 2 represents the same data for the second part of the survey (simplification of numerical expressions for a person who can not understand decimals) and Table 3 for the third part (free simplification of numerical expressions for a person with poor numeracy).

In the three parts of the survey, the percentage of simplifications that use hedges is slightly higher than that of those not using hedges especially in the second and third part of the survey. Adapting original numerical expressions by inserting hedges accounts for more than the 50% of cases. This reinforces our assumption that simplifications involving loss of precision may be better understood if an appropriate hedge is used.

4.1 Analysis of the Use of Hedges in the Simplified Numerical Expressions

In order to test hypothesis H1 (the use of hedges in the simplified numerical expression is influenced by the simplification strategy selected), we carried out a series of two sample *t-tests* where statistical significance was adjusted for multiple comparisons by using the *Bonferroni correction*. Results are presented in Table 4. When considering the entire survey (*Whole* column), there is no significant difference in the use of hedges in fractions and percentages. When analyzing the survey by parts we find similar results. There is no significant difference in the use of hedges in any strategy in the second (no decimals) and the third (free simplification) parts of

Num. Exp.		Frequency (%)	Error (%)	Hedge (%)
more than 1%	Fractions	18	0	67
	Ratios	6	0	100
	Percentages	18	17	50
2%	Fractions	6	0	50
	Ratios	18	-1	17
	Percentages	12	0	0
16.8%	Fractions	26	1	67
	Ratios	65	5	45
	Percentages	9	-3	0
27%	Fractions	82	-4	86
	Ratios	12	8	75
	Percentages	6	6	50
at least 30%	Fractions	41	0	93
	Ratios	35	13	67
	Percentages	3	0	100
40%	Fractions	53	12	50
	Ratios	29	0	10
	Percentages	6	0	0
56%	Fractions	82	-13	82
	Ratios			
	Percentages	6	-5	50
63%	Fractions	74	-3	84
	Ratios	24	0	75
	Percentages	3	0	0
75%	Fractions	32	0	0
	Ratios	29	0	0
	Percentages			
97.2%	Fractions	3	0	0
	Ratios	38	-8	23
	Percentages	18	1	50
98%	Fractions	6	0	0
	Ratios	12	0	0
	Percentages	3	0	0
Average	Fractions	39	-1	53
	Ratios	24	2	41
	Percentages	7	1	30

Table 1: Analysis of the data for 34 participants from the first part of the survey (simplifications intended for people who do not understand percentages). All values are percentages. The first column represents the frequencies of use for each simplification strategy. The second column shows the error as the loss of precision involved in the simplification. And the last column displays the use of hedges in the simplifications.

Num. Exp.		Frequency (%)	Error (%)	Hedge (%)
0.6%	Fractions	6	25	50
	Ratios	9	22	33
	Percentages	47	21	100
2.8%	Fractions	3	-29	0
	Ratios	24	6	63
	Percentages	47	7	63
6.1%	Fractions			
	Ratios	18	-4	50
	Percentages	50	-3	82
7.5%	Fractions	12	9	75
	Ratios	12	-10	0
	Percentages	50	7	41
15.5%	Fractions	15	-1	80
	Ratios	12	6	50
	Percentages	44	2	33
25.9%	Fractions	15	-3	100
	Ratios	12	-3	75
	Percentages	38	5	62
29.1%	Fractions	3	0	0
	Ratios	15	3	60
	Percentages	50	2	71
35.4%	Fractions	12	-5	100
	Ratios	15	-4	60
	Percentages	41	-1	71
50.8%	Fractions	44	-2	93
	Ratios	3	0	0
	Percentages	21	0	43
73.9%	Fractions	44	1	93
	Ratios	6	1	50
	Percentages	18	0	50
87.8%	Fractions	3	0	0
	Ratios	15	-1	60
	Percentages	47	1	88
96.9%	Fractions	3	0	0
	Ratios	12	-2	75
	Percentages	29	0	80
96.9%	Fractions	6	0	50
	Ratios	18	-1	67
	Percentages	21	0	86
97.2%	Fractions	3	0	0
	Ratios	18	-1	67
	Percentages	41	0	93
97.2%	Fractions	3	0	0
	Ratios	18	-1	83
	Percentages	32	0	91
98.2%	Fractions	3	0	0
	Ratios	15	-2	40
	Percentages	44	0	67
Average	Fractions	11	0	43
	Ratios	14	1	52
	Percentages	39	2	70

Table 2: Analysis of the data for 34 participants from the second part of the survey (simplifications intended for people who do not understand decimals). All values are percentages. The first column represents the frequencies of use for each simplification strategy. The second column shows the error as the loss of precision involved in the simplification. And the last column displays the use of hedges in the simplifications.

the survey, but in the first part (no percentages) we find significant difference between fractions and ratios ($p<0.0006$). These results do not support the hypothesis, as there is not a direct relation between the use of hedges and the selected strategy.

We performed another *t-test* adjusted by using the *Bonferroni correction* on the simplification strategies and central and peripheral values to test hypothesis H2 (the use of hedges to simplify the numerical expression is influenced by the value of the proportion, with values in the central range (say 0.2 to 0.8) and values at the extreme ranges (say 0.0-0.2 and 0.8-1.0) having a different use of hedges). In this case there is also no significant difference. The results show that the use of hedges is not influenced by central and peripheral values, rejecting our hypothesis H2 with a p-value $p=0.77$ in the worst case for the percentages strategy.

A new *t-test* adjusted by using the *Bonferroni correction* was done to test hypothesis H3 (the loss of precision allowed for the simplified numerical expression is influenced by the simplification strategy selected). Table 5 shows significant differences between each simplification strategy and each kind of simplification. In the *Whole* column we can observe that the loss of precision in fractions is significantly different to the one in ratios and percentages. In the first part (no percentages) there is a significant difference between ratios and the rest of simplification strategies. In the second part (no decimals) there is

Num. Exp.		Frequency (%)	Error (%)	Hedge (%)
0.7%	Fractions			
	Ratios	6	43	100
	Percentages	9	43	100
12%	Fractions	6	-17	100
	Ratios	21	-8	71
	Percentages	21	-17	100
26%	Fractions	41	-4	57
	Ratios	12	-4	50
	Percentages			
36%	Fractions	41	-8	86
	Ratios	9	-2	67
	Percentages			
53%	Fractions	41	-6	50
	Ratios			
	Percentages	6	-6	50
65%	Fractions	21	-5	100
	Ratios	18	-1	33
	Percentages	3	0	0
75%	Fractions	15	0	20
	Ratios	9	0	33
	Percentages	3	0	0
91%	Fractions			
	Ratios	29	-1	50
	Percentages	6	-1	50
above 97%	Fractions			
	Ratios	32	0	64
	Percentages	6	2	100
Average	Fractions	18	-7	69
	Ratios	15	3	59
	Percentages	6	3	57

Table 3: Analysis of the data for 34 participants from the third part of the survey (free simplification intended for people with poor literacy). All values are percentages. The first column represents the frequencies of use for each simplification strategy. The second column shows the error as the loss of precision involved in the simplification. And the last column displays the use of hedges in the simplifications.

no significant difference between any strategy. And in the last part (free simplification) there is only a significant difference between fractions and ratios. These results seem not to support the hypothesis, as there is not a direct relation between the use of hedges and the loss of precision in the simplified numerical expression.

For hypothesis H4 (there is some kind of correlation between the loss of precision and the use of hedges), we looked for correlations between each part of the survey and each kind of simplification strategy. We carried out a non-parametric measure of statistical dependence between the two variables (loss of precision and use of hedges) calculated by the *Spearman's rank correlation coefficient*.

In general, the results show no correlation, so there is no linear dependence between the loss of precision in the strategy and use of hedges, rejecting our hypothesis. For example, there are cases with a weak correlation (e.g. in the second part of the survey for fractions with r=0.49, N=17 and p=0.03), and cases where there is a strong correlation (e.g. in the third part of the survey, with r=1, N=18 and p<.0001).

Finally, when we analyzed hypothesis H5 (when writers choose numerical expressions for readers with low numeracy, they do not tend to use hedges if they are not losing precision), we worked with each part of the survey to study the cases where the loss of precision is zero and what is the tendency of use of hedges.

- In the first part of the survey (simplification of numerical expressions for a person who can not understand percentages), considering our 34 participants, in a 46% of responses the loss of precision is zero, and for these cases only 11% used hedges.

- For the second part (simplification of numerical expressions for a person who can not understand decimals), considering our 34 participants, in a 16% of responses the loss of precision is zero and for these cases only 7% used hedges.

- And finally, in the last part (simplification of numerical expressions for a person with poor numeracy), considering the same participants, in a 23% of cases the loss of precision is zero in the simplification and for these cases only 6% used hedges.

With this data, it seems that we can accept hypothesis H5, that is, we found evidence for our assumption that when writers choose numerical expressions for readers with poor numeracy, they tend to use hedges when they round the original numerical expression, i.e when the loss of precision is not zero.

4.2 Original Numerical Expressions with Hedges

In our survey there were a few cases where the original numerical expression had a hedge. We have observed that if the original numerical expression has hedge almost always the simplified numerical expression contained a hedge. There is a special case, "above 97%" where we do not count the use of hedges because in this case the participants chose non-numeric options mostly and they rewrote the numerical expression with phrases like "around all".

Strategy	No Pct.		No Dec.	Free Simp.	Whole	
Fractions	A		A	A	A	
Percentages	A	B	A	A	A	
Ratios		B	A	A		B

Table 4: Results of t-test adjusted by Bonferroni correction for H1 (the use of hedges in simplified numerical expressions is influenced by the simplification strategy selected). Strategies which do not share a letter are significantly different.

Strategy	No Pct.		No Dec.		Free Simp.		Whole	
Fractions	A		A		A		A	
Percentages	A		A		A	B		B
Ratios		B	A			B		B

Table 5: Results of t-test adjusted by Bonferroni correction for H3 (the loss of precision allowed for the simplified numerical expression is influenced by the simplification strategy selected). Strategies which do not share a letter are significantly different.

In the remaining cases, the same hedge is nearly alway chosen to simplify the numerical expression.

4.3 Kinds of Hedges

With respect to the actual hedges used, we have identified two different possible roles of hedge ingredients in a numerical expression. In some cases, hedges are used to indicate that the actual numerical value given is an approximation to the intended value. Uses of *about* or *around* are instances of this. This kind of hedge is employed to indicate explicitly that some loss of precision has taken place during simplification. In other cases, hedges are used to indicate the direction in which the simplified value diverges from the original value. Uses of *under* or *over* are instances of this. In some cases more than one hedge may be added to an expression to indicate both approximation and direction, or to somehow specify the precision involved in the simplification, as in *just under* or *a little less than*.

In our analysis we studied which hedges were the most frequent in each part of the survey. Only hedges with more than ten appearances in total (including simplification strategies not present in the table) have been considered in Table 6. We observed that the three parts of the survey have three hedges in common: *about, just over* and *over*. They are used in different strategies for each kind of simplification. In the second part of the survey, where simplifications of numerical expressions for a person who can not understand decimals are done, is where more hedges are used, in special for percentages strategy. In the last part of the survey, where there is more freedom to decide how simplify the original numerical expression, participants used less hedges compare to the others parts.

No Percentages			
Hedge	Fractions	Ratios	Percent.
about	15	9	0
at least	8	5	1
just over	21	1	0
more than	9	3	0
over	6	3	2
Total	59	21	3
No Decimals			
Hedges	Fractions	Ratios	Percent.
about	8	12	6
almost	4	1	8
just over	13	3	39
just under	3	2	27
nearly	7	5	24
over	7	5	9
Total	42	28	113
Free Simplification			
Hedges	Fractions	Ratios	Percent.
about	6	5	1
just over	6	0	5
more than	4	5	0
nearly	4	0	2
over	11	2	3
Total	31	12	11

Table 6: Use of the most frequent hedges in each part of the survey

5 Discussion

As can be seen in the results, the use of hedges to simplify numerical expressions can be influenced by three parameters. The first is the *kind of simplification*. Our survey was divided in three parts depending on the mathematical knowledge of the final user. The second is the *simplification strategy* for choosing mathematical form (fractions, ratios, or percentages). In our data we observed some differences in the usage of hedges with ratios and their usage with fractions and percentages (see Table 4). The last parameter is the *loss of precision* that occurs when the numerical expression is rounded. We investigated the use of hedges vs. loss of precision with different tests hoping to define some dependencies, but there was no clear correlation between them, and it was only when we tried a deeper analysis of strategies and kind of simplifications that we found some correlations such as those we presented in Section 4.1.

When asked to simplify for people who do not understand percentages, or for people with poor numeracy, the participants use different simplification strategies and sometimes they use hedges to simplify the original numerical expression. As some participants commented, not only are percentages mathematically sophisticated forms, but they may be used in sophisticated ways in the text, often for example describing rising and falling values, for which increases or decreases can themselves be described in percentages terms. Such complex relationships are likely to pose problems for people with poor numeracy even if a suitable strategy can be found for simplifying the individual percentages. In some of the examples with more than one numerical expression being compared, some of the evaluators reported a tendency to phrase them both according to a comparable base. Thus we should consider the role of context (the set of numerical expressions in a given sentence as a whole, and the meaning of the text) in establishing what simplifications must be used.

6 Conclusions and Future Work

Through a survey administered to experts on numeracy, we have collected a wide range of examples of appropriate simplifications of percentage expressions. These examples of simplified expressions give us information about the use of hedges that our participants carry out to adapt the original numerical expression to be understood by the final user. We investigated the loss of precision that occurs with each hedge and the relation between the simplification strategy and the use of hedges.

Our aim is to use this data to guide the development of a system for automatically simplifying percentages in texts. With the knowledge acquired from our study we will improve our algorithm to simplify numerical expressions. We could determinate from the simplification strategy, kind of simplification and the loss of precision allowed, which will be the best option to adapt the original numerical expression to the final user and if that option uses hedges to understand better the original numerical expression. As a part of our algorithm, we will have to look at interrater agreements for identifying appropriate hedges.

As future work, we plan to carry out another study to determine a ranking of simplification strategies from collecting a repertoire of rewriting strategies used to simplify. This data should allow us to determine whether common values are considered simpler and whether the value of the original expression influences the chosen simplification strategy. So, given a numerical expression, we could choose what simplification strategy to apply and whether to insert a hedge. We could investigate whether the value of the original proportion also influences choices, depending on its correspondence with central or peripheral values.

We have also collected a parallel corpus of numerical expressions (original vs. simplified version). This corpus will be shared with other researches so it can be used in different applications to improve the readability of text. This could be a very useful resource because simplification of percentages remains an interesting and non-trivial problem.

Acknowledgments

This research is funded by the Spanish Ministry of Education and Science (TIN2009-14659-C03-01 Project), Universidad Complutense de Madrid and Banco Santander Central Hispano (GR58/08 Research Group Grant), and the FPI grant program.

References

Stephanie Schinzing Marsiglio Ann M. Bisantz and Jessica Munch. 2005. Displaying uncertainty: Investigating the effects of display format and specificity. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(4):777.

J. Carroll, G. Minnen, Y. Canning, S. Devlin, and J. Tait. 1998. Practical simplification of English newspaper text to assist aphasic readers. In *AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, Madison, Wisconsin.

Raman Chandrasekar, Christine Doran, and Bangalore Srinivas. 1996. Motivations and Methods for Text Simplification. In *COLING*, pages 1041–1044.

Paul Slovic Ellen Peters, Judith Hibbard and Nathan Dieckmann. 2007. Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs*, 26(3):741–748.

Shiv B. Mishra H, Mishra A. 2011. In praise of vagueness: malleability of vague information as a performance booster. *Psychological Science*, 22(6):733–8, April.

Paul Slovic Nathan F. Dieckmann and Ellen M. Peters. 2009. The use of narrative evidence and explicit likelihood by decisionmakers varying in numeracy. *Risk Analysis*, 29(10).

The United Nations. 1994. Normas uniformes sobre la igualdad de oportunidades para las personas con discapacidad. Technical report.

Sarah E. Petersen and Mari Ostendorf. 2007. Text Simplification for Language Learners: A Corpus Analysis. *Speech and Language Technology for Education (SLaTE)*.

Qualification and Curriculum Authority. 1999. Mathematics: the national curriculum for england. Department for Education and Employment, London.

Advaith Siddharthan. 2002. Resolving Attachment and Clause Boundary Amgiguities for Simplifying Relative Clause Constructs. In *Proceedings of the Student Research Workshop, 40th Meeting of the Association for Computacional Linguistics*.

Sandra Williams and Richard Power. 2009. Precision and mathematical form in first and subsequent mentions of numerical facts and their relation to document structure. In *Proceedings of the 12th European Workshop on Natural Language Generation*, Athens.

Sandra Williams and Ehud Reiter. 2005. Generating readable texts for readers with low basic skills. In *Proceeding of the 10th European Workshop on Natural Language Generation*, pages 140–147, Aberdeen, Scotland.

Joel Williams, Sam Clemens, Karin Oleinikova, and Karen Tarvin. 2003. The Skills for Life survey: A national needs and impact survey of literacy, numeracy and ICT skills. Technical Report Research Report 490, Department for Education and Skills.